Development of Research Data Management Service for Open Science in Japan

Yusuke Komiyama[1*]

[1*] *National Institute of Informatics, Research Organization of Information and Systems, 2-1-2 Hitotsubashi, Chiyoda Ward, Tokyo 101-8430, Japan*
Email: komiyama@nii.ac.jp

Summary. National Institute of Informatics (NII) will provide three research infrastructures for publication, discovery, and management to promote open science in Japan. In fact, to introduce to universities and research institutions there are barriers such as classical computer systems, research conventions, and laws, organizational governance. We are developing a data infrastructure for open science that combines organizational issues with the researchers' usability. In this research, we report on the system development and trial results of GakuNin RDM which is the research data management service.

Keywords. Research Data Management Service, RDM, Open Science, research integrity, data platform.

1. Introduction

National Institute of Informatics in Research Organization of Information and Systems (NII, ROIS) develops three types of research data infrastructure for open science in Japanese academia. They are research data publication, discovery, and management services. This report will describe the research and development of the data management infrastructure.

We propose the function of research data management for governance management to chief information officer (CIO) to extend the open science infrastructure to Japanese academia. For example, it is that RDM has technology such as research trail and data encryption. In addition to improving researcher's workflow in RDM development, we also need to consider the merits of organization managers at the same time.

So, we will provide a new research data management (RDM) service, and named it GakuNin RDM.

2. System Development

2.1 Core System

We adopted Open Science Framework (OSF) as base system, the open source software developed by Center for Open Science (COS) in the US as the core system of RDM [1]. However, since the original OSF is not compliant with classical research practices of Japan, the Japanese version needs to be customized. It also extends management functions for IT administrators in academic institutions.

2.2 High-speed network and Cloud connection

We are preparing to connect service and cloud provider with L2VPN with SINET which is 100 Gbps high-speed network for science in Japan [2]. SINET has more than 800 usage agencies and can connect systems with more than 20 cloud providers. GakuNin RDM is built on this network.

2.3 Academic Access Management Federation

GakuNin RDM is compatible with Academic Access Management Federation of Japan; it is called GakuNin [3]. Users of universities and research institutions will be able to log-in to the system with the ID of their institutions. Also, the user can log-in to the service provider (SP) corresponding to GakuNin federation with single sign-on.

2.4 Cloud Storage Add-on

We have developed two cloud object storage plugins used in Japanese universities. One is the public Cloud Azure Blob Storage, and the private Cloud is Open Stack Swift (API v 2). These were

not implemented in the original OSF, so we feedback to COS on a pull request.

2.5 Research Data Repository Add-on

On the other hand, we also developed a plugin for the data repository for university libraries and research libraries. The system can send data from GakuNin RDM to WEKO2 which is a famous Japanese repository software.

2.6 Data Analysis Add-on

Also, we have implemented the ability to send data directly from the RDM to the data analysis platform to make the research workflow comfortable. We have adopted JupyterHub as a data analysis platform and currently only Python, it can be extended to various programming languages [4].

3. Trial and Use-case Reports

3.1 Trial for university IT centers

We conducted the first closed alpha test of GakuNin RDM in February 2016. Especially it was an experiment for faculty of six universities IT center and one IT department of a research institute. The number of participating faculties was about ten people. This experiment was conducted specifically to ask usability evaluation and request. As a review of GakuNin RDM, faculties at IT center pointed out that add-on control for each organization, management of users, a usability of a user interface, management of research trails and logs are insufficient.

To respond to these challenges, we are currently developing functions for institutional IT administrators.

3.2 Trial for laboratory in wide disciplines

Furthermore, we conducted the second closed alpha test of GakuNin RDM in October 2017. It is under experiment at the stage when the authors are writing this manuscript. The purpose of this experiment was to know the use cases when researchers use them in the laboratory and to receive reviews from them.

4. Conclusions

We have developed a prototype of research data management service GakuNin RDM for open science in Japanese academia. It is a nationwide RDM SaaS to promote super interdisciplinary research and experimental reproducibility by data-driven science, it promotes open science. The new service will be launched in 2020 as a business operation of NII.

In the first closed alpha test in fiscal 2016, faculty members of IT center of six universities and one research institution evaluated the function of the service. IT experts requested the management function to the service for administrative staffs in the university, and make the guideline of Cloud utilization. In the second Closed alpha test in fiscal 2017, researcher of five universities and one institution reviewed the service. At the second Closed alpha test, we investigated collaborative research method in various disciplines and cooperation with external systems such as research instruments or analytical software.

We have expanded the functions for institutional IT administrators to GakuNin RDM so far. In the future, we plan to enrich modules for RDM services for diverse academic fields in response to requests from researchers.

References

1. Foster, E. D., et al., Open Science Framework (OSF). *J. Med. Libr. Assoc.*, 105(2), 38, 2017
2. Kurimoto, T., et al., A fully meshed backbone network for data-intensive sciences and SDN services. *in 2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN)*, 909–911, 2016
3. Yamaji, K., et al., Attribute Aggregating System for Shibboleth Based Access Management Federation. *in 2010 10th IEEE/IPSJ International Symposium on Applications and the Internet*, 281–284, 2010
4. Grüning, B. A., et al., Jupyter and Galaxy: Easing entry barriers into complex data analyses for biomedical researchers. *PLoS Comput. Biol.*, 13(5), e1005425, 2017